of a single ddNTP. The DNA polymerase occasionally incorporates a dideoxynucleotide which terminates chain extension. Because the dideoxynucleotide has no 3'-hydroxyl, the initiation point for the polymerase enzyme is lost. Polymerization produces a mixture of fragments of varied sizes, all having identical 3' termini. Fractionation of the mixture by, for example, polyacrylamide gel electrophoresis, produces a pattern which indicates the presence and position of each base in the nucleic acid. Reactions with each of the four ddNTPs allows one of ordinary skill to read an entire nucleic acid sequence from a resolved gel.

Please replace the paragraph on page 3, line 23 through page 4, line 6 with the following:

A second drawback is the poor level of discrimination between correctly hybridized, perfectly matched duplexes, and an end mismatch. In part, these drawbacks have been addressed at least to a small degree by the method of continuous stacking hybridization as reported by Khrapko *et al.* (FEBS Lett. 256:118-22, 1989). Continuous stacking hybridization is based upon the observation that when a single-stranded oligonucleotide is hybridized adjacent to a double-stranded oligonucleotide, the two duplexes are mutually stabilized as if they are positioned side-to-side due to a stacking contact between them. The stability of the interaction decreases significantly as stacking is disrupted by nucleotide displacement, gap, or terminal mismatch. Internal mismatches are presumably ignorable because their thermodynamic stability is so much less than perfect matches. Although promising, a related problem arises which is the inability to distinguish between weak, but correct duplex formation, and simple background such as non-specific adsorption of probes to the underlying support matrix.

Please replace the paragraph on page 4, lines 18-25 with the following:

A final drawback is the possibility that certain probes will have anomalous behavior and, for one reason or another, be recalcitrant to hybridization under whatever standard sets of conditions are ultimately used. A simple example of this is the difficulty in finding matching conditions for probes rich in G/C

content. A more complex example could be sequences with a high propensity to form triple helices. The only way to rigorously explore these possibilities is to carry out extensive hybridization studies with all possible oligonucleotides of length n, under the particular format and conditions chosen. This is clearly impractical if many sets of conditions are involved.

Please replace the paragraph on page 3, line 23 through page 4, line 6 with the following:

Among the early publications which appeared discussing sequencing by hybridization, E. M. Southern (PCT application no. WO 89/10977, published Nov. 16, 1989; which is hereby specifically incorporated by reference), described methods whereby unknown, or target, nucleic acids are labeled, hybridized to a set of nucleotides of chosen length on a solid support, and the nucleotide sequence of the target determined, at least partially, from knowledge of the sequence of the bound fragments and the pattern of hybridization observed. Although promising, as a practical matter, this method has numerous drawbacks. Probes are entirely single-stranded and binding stability is dependant upon the size of the duplex. However, every additional nucleotide of the probe necessarily increases the size of the array by four fold, creating a dichotomy which severely restricts its plausible use. Further, there is an inability to deal with branch point ambiguities or secondary structure of the target, and hybridization conditions will have to be tailored or in some way accounted for for each binding event.

Please replace the paragraph on page 5, lines 10-14 with the following:

R. Drmanac *et al.* (U.S. Pat. No. 5,202,231; which is specifically incorporated by reference) is directed to methods for sequencing by hybridization using sets of oligonucleotide probes with random sequences. These probes, although useful, suffer from some of the same drawbacks as the methodology of Southern (1989), and like Southern, fail to recognize the advantages of stacking interactions.

Please replace the paragraph on page 6, lines 3-9 with the following:

One embodiment of the invention is directed to arrays of 4^R different nucleic acid probes wherein each probe comprises a double-stranded portion of length D, a terminal single-stranded portion of length S, and a random nucleotide sequence within the single-stranded portion of length R. These arrays may be bound to solid supports and are useful for determining the nucleotide sequence of unknown nucleic acids and for the detection, identification and purification of target nucleic acids in biological samples.

Please replace the paragraph on page 6, lines 10-15 with the following:

Another embodiment of the invention is directed to methods for creating arrays of probes comprising the steps of synthesizing a first set of nucleic acids each comprising a constant sequence of length C at the 3'-terminus, and a random sequence of length R at the 5'-terminus, synthesizing a second set of nucleic acids each comprising a sequence complementary to the constant sequence of the first nucleic acid, and hybridizing the first set with the second set to form the array.

Please replace the paragraph on page 6, lines 16-24 with the following:

Another embodiment of the invention is directed to methods for creating arrays of probes comprising the steps of synthesizing a set of nucleic acids each containing a random internal sequence of length R flanked by the cleavage sites of a restriction enzyme, synthesizing a set of primers each complementary to a non-random sequence of the nucleic acid, hybridizing the two sets together to form hybrids, extending the sequence of the primer by polymerization using the nucleic acid as a template, and cleaving the hybrids with the restriction enzyme to form an array of probes with a double-stranded portion and a single-stranded portion and with the random sequence within the single-stranded portion.

Please replace the paragraph on page 7, line 16 through page 8, line 7 with the following:

Another embodiment of the invention is directed to nucleic acid probes and methods for creating nucleic acid probes comprising the steps of

synthesizing a plurality of single-stranded first nucleic acids and a plurality of longer single-stranded second nucleic acids wherein each second nucleic acid comprises a random terminal sequence and a sequence complementary to a sequence of the first nucleic acids, hybridizing the first nucleic acids to the second to form partial duplexes having a double-stranded portion and a single-stranded portion with the random sequence within the single-stranded portion, hybridizing a target nucleic acid to the partial duplexes, optionally ligating the hybridized target to the first nucleic acid of the partial duplexes, isolating the second nucleic acid from the ligated duplexes, synthesizing a plurality of third nucleic acids each complementary to the constant sequence of the second nucleic acid, and hybridizing the third nucleic acids with the isolated second nucleic acids to create the nucleic acid probe. Alternatively, after formation of the partial duplexes, the target is ligated as before and hybridized with a set of oligonucleotides comprising random sequences. These oligonucleotides are ligated to the second nucleic acid, the second nucleic acid is isolated, another plurality of first nucleic acids are synthesized, and the first nucleic acids are hybridized to the oligonucleotide ligated second nucleic acids to form the probe. Ligation allows for hybridization to be performed under a single set of hybridization conditions. Probes may be fixed to a solid support and may also contain enzyme recognition sites within their sequences.

Please replace the paragraph on page 8, lines 20-22 with the following:

FIG. 1 Energetics of stacking hybridization. Structures consist of a long target and a probe of length n. The top three samples are ordinary hybridization and the bottom three are stacking hybridization.

Please replace the paragraph on page 9, lines 28-29 with the following:

FIG. 14 A diagrammatic representation of the construction of a complementary array of master beads.

Please replace the paragraph on page 10, lines 2-11 with the following:

The present invention overcomes the problems and disadvantages associated with current strategies and designs and provides new methods and probes, new diagnostic aids and methods for using the diagnostic aids, and new

arrays and methods for creating arrays of probes to detect, identify, purify and sequence target nucleic acids. Nucleic acids of the invention include sequences of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) which may be isolated from natural sources, recombinantly produced, or artificially synthesized. Preferred embodiments of the present invention are probes synthesized using traditional chemical synthesis, using the more rapid polymerase chain reaction (PCR) technology, or using a combination of these two methods.

Please replace the paragraph on page 10, lines 12-19 with the following:

Nucleic acids of the invention further encompass polyamide nucleic acid (PNA) or any sequence of what are commonly referred to as bases joined by a chemical backbone that have the ability to base pair or hybridize with a complementary chemical structure. The bases of DNA, RNA, and PNA are purines and pyrimidines linearly linked to a chemical backbone. Common chemical backbone structures are deoxyribose phosphate and ribose phosphate. Recent studies demonstrated that a number of additional structures may also be effective, such as the polyamide backbone of PNA (P. E. Nielsen *et al.*, Sci. 254:1497-1500, 1991).

Please replace the paragraph on page 10, line 20 through page 11, line 2 with the following:

The purines found in both DNA and RNA are adenine and guanine, but others known to exist are xanthine, hypoxanthine, 2- and 1-diaminopurine, and other more modified bases. The pyrimidines are cytosine, which is common to both DNA and RNA, uracil found predominantly in RNA, and thymidine which occurs exclusively in DNA. Some of the more atypical pyrimidines include methylcytosine, hydroxymethyl-cytosine, methyluracil, hydroxymethyluracil, dihydroxypentyluracil, and other base modifications. These bases interact in a complementary fashion to form base-pairs, such as, for example, guanine with cytosine and adenine with thymidine. However, this invention also encompasses situations in which there is nontraditional base pairing such as Hoogsteen base pairing which has been identified in certain tRNA molecules and postulated to exist in a triple helix.

Please replace the paragraph on page 11, lines 21-26 with the following:

By way of example only, if the random portion consisted of a four nucleotide sequence (R = 4) of adenine, guanine, thymine, and cytosine, the total number of possible combinations (4^R) would be 4^4 or 256 different nucleic acid probes. If the number of nucleotides in the random sequence was five, the number of different probes within the set would be 4^5 or 1,024. This becomes a very large number indeed when considering sequences of 20 nucleotides or more.

Please replace the paragraph on page 12, line 13 through page 13, line 2 with the following:

Hybridization between complementary bases of DNA, RNA, PNA, or combinations of DNA, RNA and PNA, occurs under a wide variety of conditions such as variations in temperature, salt concentration, electrostatic strength, and buffer composition. Examples of these conditions and methods for applying them are described in Nucleic Acid Hybridization: A Practical Approach (B. D. Hames and S. J. Higgins, editors, IRL Press, 1985), which is herein specifically incorporated by reference. It is preferred that hybridization takes place between about 0°C and about 70°C, for periods of from about 5 minutes to hours, depending on the nature of the sequence to be hybridized and its length. For example, typical hybridization conditions for a mixture of two 20-mers is to bring the mixture to 68°C and let it cool to room temperature (22°C) for five minutes or at very low temperatures such as 2°C in 2 microliters. It is also preferred that hybridization between nucleic acids be facilitated using buffers such as saline, Tris-EDTA (TE), Tris-HCl and other aqueous solutions, certain reagents and chemicals. Preferred examples of these reagents include single-stranded binding proteins such as Rec A protein, T4 gene 32 protein, *E. coli* single-stranded binding protein, and major or minor nucleic acid groove binding proteins. Preferred examples of other reagents and chemicals include divalent ions, polyvalent ions, and intercalating substances such as ethidium bromide, actinomycin D, psoralen, and angelicin.

Please replace the paragraph on page 14, line 33 through page 15, line 9 with the following:

Another embodiment of the invention is directed to methods for determining a sequence of a nucleic acid comprising the steps of labeling the nucleic acid with a first detectable label at a terminal site, labeling the nucleic acid with a second detectable label at an internal site, identifying the nucleotide sequences of portions of the nucleic acid, determining the relationship of the nucleotide sequence portions to the nucleic acid by comparing the first detectable label and the second detectable label, and determining the nucleotide sequence of the nucleic acid. Fragments of target nucleic acids labeled both terminally and internally can be distinguished based on the relative amounts of each label within respective fragments. Fragments of a target nucleic acid terminally labeled with a first detectable label will have the same amount of label as fragments which include the labeled terminus. However, these fragments will have variable amounts of the internal label directly proportional to their size and distance for the terminus. By comparing the relative amount of the first label to the relative amount of the second label in each fragment, one of ordinary skill is able to determine the position of the fragment or the position of the nucleotide sequence of that fragment within the whole nucleic acid.

Please replace the paragraph on page 15, line 10 through page 16, line 2 with the following:

Another embodiment of the invention is directed to methods for determining a nucleotide sequence by hybridization comprising the steps of creating a set of nucleic acid probes wherein each probe has a double-stranded portion, a single-stranded portion, and a random sequence within the single-stranded portion which is determinable, hybridizing a nucleic acid target which is at least partly single-stranded to the set, ligating the hybridized target to the probe, and determining the nucleic sequence of the target which is hybridized to the single-stranded portion of any probe. This embodiment adds a step wherein the hybridized target is ligated to the probe. Ligation of the target nucleic acid to the complementary probe increases fidelity of hybridization and

allows for incorrectly hybridized target to be easily washed from correctly hybridized target (FIG. 11). More importantly, the addition of a ligation step allows for hybridization to be performed under a single set of hybridization conditions. For example, hybridization temperature is preferably between about 22°-37°C, the salt concentration useful is preferably between about 0.05-0.5M, and the period of hybridization is between about 1-14 hours. This is not possible using the methodologies of the current procedures which do not employ a ligation step and represents a very substantial improvement. Ligation can be accomplished using a eukaryotic-derived or a prokaryotic-derived ligase. Preferred is T4 DNA or RNA ligase. Methods for use of these and other nucleic acid modifying enzymes are described in Current Protocols in Molecular Biology (F. M. Ausubel *et al.*, editors, John Wiley & Sons, 1989), which is herein specifically incorporated by reference.

Please replace the paragraph on page 16, lines 3-11 with the following:

There are a number of distinct advantages to the incorporation of a ligation step. First and foremost is that one can use identical hybridization conditions for hybridization. Variation of hybridization conditions due to base composition are no longer relevant as nucleic acids with high A/T or G/C content ligate with equal efficiency. Consequently, discrimination is very high between matches and mis-matches, much higher than has been achieved using other methodologies such as Southern (1989) wherein the effects of G/C content were only somewhat neutralized in high concentrations of quaternary or tertiary amines (e.g., 3M tetramethyl ammonium chloride in Drmanac *et al.*, 1993).

Please replace the paragraph on page 16, line 12 through page 17, line 2 with the following:

Another embodiment of the invention is directed to methods for determining a nucleotide sequence by hybridization which comprises the steps of creating a set of nucleic acid probes wherein each probe has a double-stranded portion, a single-stranded portion, and a random sequence within the single-stranded portion which is determinable, hybridizing a target

nucleic acid which is at least partly single-stranded to the set of nucleic acid probes, enzymatically extending a strand of the probe using the hybridized target as a template, and determining the nucleotide sequence of the single-stranded portion of the target nucleic acid. This embodiment of the invention is similar to the previous embodiment, as broadly described herein, and includes all of the aspects and advantages described therein. An alternative embodiment also includes a step wherein hybridized target is ligated to the probe. Ligation increases the fidelity of the hybridization and allows for a more stringent wash step wherein incorrectly hybridized, unligated target can be removed and further allows for a single set of hybridization conditions to be employed. Most nonligation techniques including Southern (1989), Drmanac *et al.* (1993), and Khrapko *et al.* (1989 and 1991), are only accurate, and only marginally so, when hybridizations are performed under optimal conditions which vary with the G/C content of each interaction. Preferable conditions comprise a hybridization temperature of between about 22°-37°C, a salt concentration of between about 0.05-0.5M, and a hybridization period of between about 1-14 hours.

Please replace the paragraph on page 17, lines 10-19 with the following:

Hybridized probes may also be enzymatically extended a predetermined length. For example, reaction condition can be established wherein a single dNTP or ddNTP is utilized as substrate. Only hybridized probes wherein the first nucleotide to be incorporated is complementary to the target sequence will be extended, thus, providing additional hybridization fidelity and additional information regarding the nucleotide sequence of the target. Sanger (1977) or Maxam and Gilbert (1977) sequencing can be performed which would provide further target sequence data. Alternatively, hybridization of target to probe can produce 3′ extensions of target nucleic acids. Hybridized probes can be extended using nucleoside biphosphate substrates or short sequences which are ligated to the 5′ terminus.

Please replace the paragraph on page 17, line 20 through page 18, line 4 with the following:

Another embodiment of the invention is directed to a method for determining a nucleotide sequence of a target by hybridization comprising the steps of creating a set of nucleic acid probes wherein each probe has a double-stranded portion, a single-stranded portion, and a random nucleotide sequence within the single-stranded portion which is determinable, cleaving a plurality of nucleic acid targets to form fragments of various lengths which are at least partly single-stranded, hybridizing the single-stranded region of the fragments with the single-stranded region of the probes, identifying the nucleotide sequences of the hybridized portions of the fragments, and comparing the identified nucleotide sequences to determine the nucleotide sequence of the target. An alternative embodiment includes a further step wherein the hybridized fragments are ligated to the probes prior to identifying the nucleotide sequences of the hybridized portions of the fragments. As described herein, the addition of a ligation step allows for hybridizations to be performed under a single set of hybridization conditions.

Please replace the paragraph on page 19, lines 8-16 with the following:

Another embodiment of the invention is directed to a method wherein the target nucleic acid has a first detectable label at a terminal site and a second detectable label at an internal site. The labels may be the same type of label or of different types as long as each can be discriminated, preferably by the same detection method. It is preferred that the first and second detectable labels are chromatic or fluorescent chemicals or molecules which are detectable by mass spectrometry. Using a double-labeling method coupled with analysis by mass spectrometry provides a very rapid and accurate sequencing methodology that can be incorporated in sequencing by hybridization and lends itself very well to automation and computer control.

Please replace the paragraph on page 19, lines 17-27 with the following:

Another embodiment of the invention is directed to methods for creating a nucleic acid probe comprising the steps of synthesizing a plurality of

single-stranded first nucleic acids and an array of longer single-stranded second nucleic acids complementary to the first nucleic acid with a random terminal nucleotide sequence, hybridizing the first nucleic acids to the second nucleic acids to form hybrids having a double-stranded portion and a single-stranded portion with the random nucleotide sequence within the single-stranded portion, hybridizing a single-stranded nucleic acid target to the hybrids, ligating the hybridized target to the first nucleic acid of the hybrid, isolating the second nucleic acid, and hybridizing the first nucleic acid of step with the isolated second nucleic acid to form a nucleic acid probe. Probes created in this manner are referred to herein as customized probes.

Please replace the paragraph on page 20, line 28 through page 21, line 8 with the following:

Another embodiment of the invention is directed to methods for creating probe arrays comprising the steps of synthesizing a first set of nucleic acids each comprising a constant sequence of length C at a 3′-terminus and a random sequence of length R at a 5′-terminus, synthesizing a second set of nucleic acids each comprising a sequence complementary to the constant sequence of each of the first nucleic acids, and hybridizing the first set with the second set to create the array. Preferably, the nucleic acids of the first set are each between about 15- 30 nucleotides in length and the nucleic acids of the second set are each between about 10-25 nucleotides in length. Also preferable is that C is between about 7-20 nucleotides and R is between about 3-10 nucleotides.

Please replace the paragraph on page 21, line 22 through page 22, line 2 with the following:

Alternatively, probe arrays may also be made which are single-stranded. These arrays are created, preferably on a solid support, basically as described, by synthesizing an array of nucleic acids each comprising a constant sequence of length C at a 3′-terminus and a random sequence of length R at a 5′-terminus, and fixing the array to a first solid support. Arrays created in this manner can be quickly and easily transformed into double-stranded arrays by the synthesis and hybridization of a set of nucleic acids with a sequence complementary to

the constant sequence of the replicated array to create a double-stranded replicated array. However, in their present form, single-stranded arrays are very valuable as templates for replication of the array.

Please replace the paragraph on page 22, lines 3-18 with the following:

Due to the very large numbers of probes which comprise most useful arrays, there is a great deal of time spent in simply creating the array. It requires many hours of nucleic acid synthesis to create each member of the array and many hours of manipulations to place the array in an organized fashion onto any solid support such as those described previously. Once the master array is created, replicated arrays, or slaves, can be quickly and easily created by the methods of the invention which take advantage of the speed and accuracy of nucleic acid polymerases. Basically, methods for replicating an array of single-stranded probes on a solid support comprise the steps of synthesizing an array of nucleic acids each comprising a constant sequence of length C at a 3'-terminus and a random sequence of length R at a 5'-terminus, fixing the array to a first solid support, synthesizing a set of nucleic acids each comprising a sequence complementary to the constant sequence, hybridizing the nucleic acids of the set with the array, enzymatically extending the nucleic acids of the set using the random sequences of the array as templates, denaturing the set of extended nucleic acids, and fixing the denatured nucleic acids of the set to a second solid support to create the replicated array of single-stranded probes.

Please replace the paragraph on page 22, line 26 through page 23, line 11 with the following:

Another embodiment of the invention is directed to methods for creating arrays of probes comprising the steps of synthesizing an array of single-stranded nucleic acids each containing a constant sequence at the 3'-terminus, another constant sequence at the 5'-terminus, and a random internal sequence of length R flanked by the cleavage site(s) of a restriction enzyme (on one or both sides), synthesizing an array of primers each complementary to a portion of the constant sequence of the 3'-terminus, hybridizing the two arrays together to

form hybrids, extending the sequence of each primer by polymerization using a sequence of the nucleic acid as a template, and cleaving the extended hybrids with the restriction enzyme to form an array of probes with a double-stranded portion at one terminus, and a single-stranded portion containing the random sequence at the opposite terminus. Preferably, the nucleic acids are each between about 10-50 nucleotides in length and R is between about 3-5 nucleotides in length. Any of the restriction enzymes which produce a 3'- or 5'-overhang after cleavage are suitable for use to make the array. Some of the restriction enzymes which are useful in this regard, and their recognition sequences, are depicted in Table 1.

Please replace the paragraph on page 24, lines 5-10 with the following:

Also preferred is that the array be fixed to a solid support such as a plastic, ceramic, metal, resin, polymer, gel, film, membrane or chip. Fixation can be accomplished by conjugating the reagents for synthesis with a specific binding protein or other similar substance and coating the surface of the support with the binding counterpart (e. g. biotin/streptavidin, F_c /protein A, and nucleic acid/nucleic acid binding protein).

Please replace the paragraph on page 24, lines 11-28 with the following:

Alternatively, another similar method for creating an array of probes comprising the steps of synthesizing an array of single-stranded nucleic acids each containing a constant sequence at the 3'- terminus, another constant sequence at the 5'-terminus, and a random internal sequence of length R flanked by the cleavage site(s) of a restriction enzyme (on one or both sides), synthesizing an array of primers with a sequence complementary to the constant sequence at the 3'-terminus, hybridizing the two arrays together to form hybrids, enzymatically extending the primers using the nucleic acids as templates to form full-length hybrids, cloning the full-length hybrids into vectors such as plasmids or phage, cloning the plasmids into competent bacteria or phage, reisolating the cloned plasmid DNA, amplifying the cloned sequences by multiple polymerase chain reactions, and cleaving the amplified sequences with the restriction enzyme to form the array of probes with a double-stranded

portion at one terminus and a single-stranded portion containing the random sequence at the opposite terminus. Using this method the array of probes may have 5'- or 3'-overhangs depending on the cleavage specificity of the restriction enzyme (e.g. Table 1). The array of probes may be fixed to a solid support such as a plastic, ceramic, metal, resin, polymer, film, gel, membranes and chip. Preferably, during PCR amplification, the reagent primers are conjugated with biotin which facilitates eventual binding to a streptavidin-coated surface.

Please replace the paragraph on page 25, lines 11-21 with the following:

Especially useful are diagnostic aids comprising probe arrays. These arrays can make the detection, identification, and sequencing of nucleic acids from biological samples exceptionally rapid and allows one to obtain multiple pieces of information from a single sample after performing a single test. Methods for detecting and/or identifying a target nucleic acid in a biological sample comprise the steps of creating an array of probes fixed to a solid support as described herein, labeling the nucleic acid of the biological sample with a detectable label, hybridizing the labeled nucleic acid to the array and detecting the sequence of the nucleic acid from a binding pattern of the label on the array. These methods for creating probe arrays and for rapidly and efficiently replicating those arrays, such as for diagnostic aids, makes the manufacture and commercial application of large numbers of arrays a possibility.

Please replace the paragraph on page 26, line 14 through page 27, line 5 with the following:

Another embodiment of the invention is directed to methods for creating a nucleic acid probe comprising the steps of synthesizing a plurality of single-stranded first nucleic acids and an array of longer single-stranded second nucleic acids complementary to the first nucleic acid with a random terminal nucleotide sequence, hybridizing the first nucleic acids to the second nucleic acids to form hybrids having a double-stranded portion and a single-stranded portion with the random nucleotide sequence within the single-stranded portion, hybridizing a single-stranded nucleic acid target to the hybrids, ligating the hybridized target to the first nucleic acid of the hybrid, hybridizing the ligated hybrid with an array

of oligonucleotides with random nucleotide sequences, ligating the hybridized oligonucleotide to the second nucleic acid of the ligated hybrid, isolating the second nucleic acid, and hybridizing another first nucleic acid with the isolated second nucleic acid to form a nucleic acid probe. Preferred is that the first nucleic acid is about 15-25 nucleotides in length, that the second nucleic acid is about 20-30 nucleotides in length, that the constant portion contain an enzyme recognition site, and that the oligonucleotides are each about 4-20 nucleotides in length. Probes may be fixed to a solid support such as a plastic, ceramic, a metal, a resin, a gel, or a membrane. It is preferred that the solid support be a two-dimensional or three-dimensional matrix with multiple probe binding sites such as a hybridization chip. Nucleic acid probes created by the method of the present invention are useful in a diagnostic aid to screen a biological sample for genetic variations of nucleic acid sequences therein.

Please replace the paragraph on page 27, lines 6-25 with the following:

Another embodiment of the invention is directed to a method for creating a nucleic acid probe comprising the steps of (a) synthesizing a plurality of single-stranded first nucleic acids and a set of longer single-stranded second nucleic acids complementary to the first nucleic acids with a random terminal nucleotide sequence, (b) hybridizing the first nucleic acids to the second nucleic acids to form hybrids having a double-stranded portion and a single-stranded portion with the random nucleotide sequence in the single-stranded portion, (c) hybridizing a single-stranded nucleic acid target to the hybrids, (d) ligating the hybridized target to the first nucleic acid of the hybrid, (e) enzymatically extending the second nucleic acid using the target as a template, (f) isolating the extended second nucleic acid, and (g) hybridizing the first nucleic acid of step (a) with the isolated second nucleic acid to form a nucleic acid probe. It is preferred that the first nucleic acid is about 15-25 nucleotides in length, that the second nucleic acid is about 20-30 nucleotides in length, and that the double-stranded portion contain an enzyme recognition site. It is also preferred that the probe be fixed to a solid support, such as a plastic, ceramic, a metal, a resin, a gel, or a membrane. A preferred solid support is a two-dimensional or

three-dimensional matrix with multiple probe binding sites, such as a hybridization chip. A further embodiment of the present invention is a diagnostic aid comprising the created nucleic acid probe and a method for using the diagnostic aid to screen a biological sample as herein described.

Please replace the paragraph on page 29, lines 1-7 with the following:

A protein fusion between streptavidin and metallothionein was recently constructed (T. Sano *et al.*, Proc. Natl. Acad. Sci. USA, 1992). Both partners in this protein fusion are fully active and these streptavidin-biotin interactions are being used to develop new methods for purification of DNA, including triplex-mediated capture of duplex DNA on magnetic microbeads (T. Ito *et al.*, Proc. Natl. Acad. Sci. USA 89:495-98, 1992) and affinity capture electrophoresis of DNA in agarose (T. Ito *et al.*, G.A.T.A., 1992).

Please replace the paragraph on page 31, lines 4-20 with the following:

Preparation of model arrays. Following the scheme shown in FIG. 2, in a single synthesis, all 1024 possible single-stranded probes with a constant 18 base stalk followed by a variable 5 base extension can be created. The 18 base extension is designed to contain two restriction enzyme cutting sites. *Hga I* generates a 5 base, 5′ overhang consisting of the variable bases N_5. *Not I* generates a 4 base, 5′ overhang at the constant end of the oligonucleotide. The synthetic 23-mer mixture will be hybridized with a complementary 18-mer to form a duplex which can then be enzymatically extended to form all 1024, 23-mer duplexes. These can be cloned by, for example, blunt end ligation, into a plasmid which lacks *Not I* sites. Colonies containing the cloned 23-base insert can be selected. Each should be a clone of one unique sequence. DNA minipreps can be cut at the constant end of the stalk, filled in with biotinylated pyrimidines, then cut at the variable end of the stalk to generate the 5 base 5′ overhang. The resulting nucleic acid can be fractionated by Qiagen columns (nucleic acid purification columns) to discard the high molecular weight material, and the nucleic acid probe will then be attached to a streptavidin-coated surface. This procedure could easily be automated in a Beckman Biomec or equivalent chemical robot to produce many identical arrays of probes.

Please replace the paragraph on page 32, line 26 through page 33, line 8 with the following:

Moderately dense arrays can be made using a typical x-y robot to spot the biotinylated compounds individually onto a streptavidin-coated surface. Using such robots, it is possible to make arrays of 2×10^4 samples in 100 to 400 cm^2 of nominal surface. The array should preferably fit in 10 cm^2, but even if forced, for unforeseen technical reasons, to compromise on an array ten times or even 50 times less dense, it will be quite suitable for testing the principles of and many of the variations on positional SBH. Commercially available streptavidin-coated beads can be adhered permanently to plastics like polystyrene, by exposing the plastic first to a brief treatment with an organic solvent like triethylamine. The resulting plastic surfaces have enormously high biotin binding capacity because of the very high surface area that results. This will suffice for radioactively labeled samples.

Please replace the paragraph on page 33, line 18 through page 34, line 2 with the following:

In certain experiments, the need for attaching oligonucleotides to surfaces may be circumvented altogether, and oligonucleotides attached to streptavidin-coated magnetic microbeads used as already done in pilot experiments. The beads can be manipulated in microtitre plates. A magnetic separator suitable for such plates can be used including the newly available compressed plates. For example, the 18 by 24 well plates (Genetix, Ltd.; USA Scientific Plastics) would allow containment of the entire array in 3 plates; this formate is well handled by existing chemical robots. It is preferable to use the more compressed 36 by 48 well format, so that the entire array would fit on a single plate. The advantages of this approach for all the experiments are that any potential complexities from surface effects can be avoided, and already-existing liquid handling, thermal control, and imaging methods can be used for all the experiments. Thus, this allows the characterization of many of the features of positional SBH before having to invest the time and effort in fabricating instruments, tools and chips.

Please replace the paragraph on page 35, lines 12-16 with the following:

There are a number of advantages to a ligation step. Physical specificity is supplanted by enzymatic specificity. Focusing on the 3′ end of the target nucleic also minimizes problems arising from stable secondary structures in the target DNA. As shown in FIG. 3B, ligation can be used to enhance the fidelity of detecting the 5′-terminal sequence of a target DNA.

Please replace the paragraph on page 35, line 17 through page 36, line 5 with the following:

DNA ligases are also used to covalently attach hybridized target DNA to the correct immobilized oligonucleotide probe. Several tests of the feasibility of the ligation scheme are shown in FIG. 3. Biotinylated probes were attached to streptavidin-coated magnetic microbeads, and annealed with a shorter, complementary, constant sequence to produce duplexes with 5 or 6 base single-stranded overhangs. One set of actual sequences used is shown in Example 14. ^{32}P-end labeled targets were allowed to hybridize to the probes. Free targets were removed by capturing the beads with a magnetic separator. DNA ligase was added and ligation was allowed to proceed at various salt concentrations. The samples were washed at room temperature, again manipulating the immobilized compounds with a magnetic separator. This should remove non-ligated material. Finally, samples were incubated at a temperature above the T_m of the duplexes, and eluted single-strand was retained after the remainder of the samples were removed by magnetic separation. The eluate at this point should consist of the ligated material. The fraction of ligation was estimated as the amount of ^{32}P recovered in the high temperature wash versus the amount recovered in both the high and low temperature washes. Results obtained are shown in FIG. 13. It is apparent that salt conditions can be found where the ligation proceeds efficiently with perfectly matched 5 or 6 base overhangs, but not with G-T mismatches.

Please replace the paragraph on page 36, lines 6-20 with the following:

The results of a more extensive set of similar experiments are shown in Tables 2-4. Table 2 looks at the effect of the position of the mismatch and

Table 3 examines the effect of base composition on the relative discrimination of perfect matches verses weakly destabilizing mismatches. These data demonstrate that: (1) effective discrimination between perfect matches and single mismatches occurs with all five base overhangs tested; (2) there is little if any effect of base composition on the amount of ligation seen or the effectiveness of match/mismatch discrimination. Thus, the serious problems of dealing with base composition effects on stability seen in ordinary SBH do not appear to be a problem for positional SBH; and (3) the worst mismatch position is, as expected, the one distal from the phosphodiester bond formed in the ligation reaction. However, any mismatches that survive in this position will be eliminated by a polymerase extension reaction, such as described herein, provided that polymerase is used, like sequenase version 2, that has no 3'-endonuclease activity or terminal transferase activity; and (4) gel electrophoresis analysis has confirmed that the putative ligation products seen in these tests are indeed the actual products synthesized.

Please replace the paragraph on page 40, lines 1-14 with the following:

The third approach to making nested samples is to use variants on plus/minus sequencing. For example, one can make a very even DNA sequencing ladder by using Sanger sequencing with a dideoxy-pppN terminator. This does not produce a ligatable end. However, it can be replaced with a ligatable end while still on the original template, by first removing the ddpppN with the 3' editing-exonuclease activity of DNA polymerase I in the absence of the one particular base at the end. Note that this accomplishes two things for the price of one. Not only does it generate a ladder with a ligatable end, but because one can pre-determine the identity of the base removed, it provides an additional nucleotide of DNA sequence information. One can use single color detection in four separate reactions, or ultimately, four color detection by mixing the results of four separate reactions prior to hybridization. If this approach is successful, it is amenable to more elaborate variations combining laddering and hybridization. Note that each of these procedures combines some of the power of ladder sequencing with the parallel processing of SBH.

Please replace the paragraph on page 43, lines 5-11 with the following:

The real power of the positional information comes, not from its application to the recurrent sequences, but to its applications to surrounding unique sequences. Their order will be determined unequivocally, assuming even moderately accurate position information, and thus, the effect of the branch point will be eliminated. For example, 10% accuracy in intensity ratios for a dual labeled 200 base pair target will provide a positional accuracy of 20 base pair. This would presumably be sufficient to resolve all but the most extraordinary recurrences.

Please replace the paragraph on page 46, line 19 through page 47, line 9 with the following:

Oligonucleotide ligation after target hybridization. Stacking hybridization without ligation has been demonstrated in a simple format. Eight-mer oligonucleotides were annealed to a target and then annealed to an adjacent 5-mer to extend the readable sequence from 8 to 13 bases. This is done with small pools of 5-mers specifically chosen to resolve ambiguities in sequence data that has already been determined by ordinary sequencing by hybridization using 8-mers alone. The method appears to work quite well, but it is cumbersome because a custom pool of 5-mers must be created to deal with each particular situation. In contrast, the approach taken herein (FIG. 9), after ligation of the target to the probe, is to ligate a mixture of 5-mers arranged in polychromatically labeled orthogonal pools. For example, using 5-mers of the form pATGCAp or pATGCddA, only a single ligation event will occur with each probe-target complex. These would be 3' labeled to avoid interference with the ligase. Only ten pools are required for a binary sieve analysis of 5-mers. In reality it would make sense to use many more, say 16, to introduce redundancy. If only four colors are available, those would require four successive hybridizations. For example, sixteen colors would allow a single hybridization. But the result of this scheme is that one reads ten bases per site in the array, equivalent to the use of 4^{10} probes, but one only has to make 2×4^5 probes. The gain in efficiency in this scheme is a factor of 500 over conventional sequencing by hybridization.

Please replace the sentence on page 49, line 25 with the following:

6 bp overlap, perfect match:

Please replace the paragraph on page 49, lines 1-13 with the following:

The biotinylated double-stranded probe was prepared in TE buffer by annealing the complementary single strands together at 68°C for five minutes followed by slow cooling to room temperature. A five-fold excess of monodisperse, polystyrene-coated magnetic beads (Dynal) coated with streptavidin was added to the double-stranded probe, which was then incubated with agitation at room temperature for 30 minutes. After ligation, the samples were subjected to two cold (4°C) washes followed by one hot (90°C) wash in TE buffer (FIG. 12). The ratio of ^{32}P in the hot supernatant to the total amount of ^{32}P was determined (FIG. 13). At high NaCl concentrations, mismatched target sequences were either not annealed or were removed in the cold washes. Under the same conditions, the matched target sequences were annealed and ligated to the probe. The final hot wash removed the non-biotinylated probe oligonucleotide. This oligonucleotide contained the labeled target if the target had been ligated to the probe.

Please replace the paragraph on page 49, lines 1-13 with the following:

Compensating for variations in base composition. A major problem in all suggested implementations of SBH is the rather marked dependence of T_m on base composition, and, at least in some cases, on base sequence. The use of unusual salts like tetramethyl ammonium halides or betaines (W. A. Rees *et al.*, Biochemistry 32:137-44, 1993) offers one approach to minimizing these varieties. Alternatively, base analogs like 2,6-diamino purine and 5-bromo U can be used instead of A and T, respectively, to increase the stability of A-T base pairs, and derivatives like 7-deazaG can be used to decrease the stability of G-C base pairs. The initial experiments shown in Table 2 indicate that the use of enzymes will eliminate many of the complications due to base sequences. This gives the approach a very significant advantage over non-enzymatic methods which require different conditions for each nucleic acid and are highly matched to GC content.

Please replace the paragraph on page 50, lines 8-26 as follows:

Data measurement, processing and interpretation. Highly automated methods for raw data handling and generation of contiguous DNA sequence from the hybridization are required for analysis of the data. Two methods of data acquisition have been used in prior SBH efforts: CCD cameras with fluorescent labels and image plate analyzers with radiolabeled samples. The latter method has the advantage that there is no problem with uniform sampling of the array. However, it is effectively limited to only two color analysis of DNA samples, by the use of ^{35}S and ^{32}P, differentially imaged through copper foil. In contrast, while CCD cameras are less well developed, the detection of many colors is possible by the use of appropriate exciting sources and filters. Four colors are available with conventional fluorescent DNA sequencing primers or terminators. More than four colors may be achievable if infra-red dyes are used. However, providing uniform excitation of the fluorescent array is not a trivial problem. Both detection schemes are used and the image plate analyzers are sure to work. The CCD camera approach will be necessary if some of the multicolor labeling schemes described in the proposal are ever to be realized. Label will be introduced into targets by standard enzymatic methods, such as the use of 5′ labeled PCR primers for 5′ labeling, internally alpha ^{32}P-labeled triphosphates or fluorescent-labeled base analogs for internal labeling, and similar compounds by filling in staggered DNA ends for 3′ labeling.

Please replace the paragraph on page 50, line 27 through page 51, line 11 with the following:

Both the Molecular Dynamics image plate analyzer and the Photometrics cooled CCD camera can deal with the same TIFF 8 bit data format. Thus, software developed for either instrument can be used to handle data measured on both instruments. This will save a great deal of unnecessary duplication in data processing software. Sequence interpretation software can be developed for reading sequencing chip data and assembling it into contiguous sequence, and is already underway in Moscow, at Argonne National Laboratory, and in the private sector. Such software is generally available in the interested user

community. The most useful examples of this software can be customized to fit the particularly special needs of this approach including polychromatic detection, incorporation of positional information, and pooling schemes. Specific software developments for constructing and decoding the orthogonal pools of samples that may ultimately be used are being developed because these procedures are also needed for enhanced physical mapping methods.

Please replace the paragraph on page 52, lines 1-4 with the following:

Tubes were placed in the Dynal MPC apparatus and the supernatant removed. Unbound streptavidin sites were sealed with 5 μl of 200 μM free biotin in water. The beads were washed several times with 80 μl TE. These beads can be stored in this state at 4°C for several weeks.

Please replace the paragraph on page 52, lines 14-22 with the following:

Polymerase I extension was performed on each tube of DNA in a total of 13 μl as follows (see Table 7): NEB buffer concentration was 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM DTT; 33 μM d(N-N_i) TP mix; 2 μM + ^{32}P dN_i TP complementary to one of the N_i bases; and polymerase I large fragment (Klenow). In the first well was added dTTP, dCTP and dGTP, to a concentration of 33 μM. ^{32}P-dATP was added to a concentration of 3 μM. dNTP stock solutions of 200 μM were pooled to lack the labelled nucleotide (i.e. Tube A contains C, G and T) adding 6.3 μl dNTP, 5 μl 200 μM dNTP, and 43 μl water. Radioactively labeled (*dNTP) stock solutions were 20 μM prepared from 2 μl [α^{32}P] dNTP, 5 μl 200 μM dNTP, and 43 μl water.

Please replace the paragraph on page 53, line 17 through page 54, line 2 with the following:

As a test of the synthesized oligo transfer, magnetic beads were suspended in 50 μl of 0.1M NaOH and incubated at room temperature for 10 minutes. The supernatant from each tube was removed and transferred to a fresh tube. Beads were incubated a second time with 50 μl of 0.1M NaOH. As many counts seemed to remain, the first set of beads were heated to 68°C in 50 μl NaOH which leached out a lot more counts. Each base was neutralized with 1M HCl followed by 50 μl of TE. Fresh Dynabeads were added to the

melted strand and incubated at room temp for 15 minutes with gentle shaking. Supernatants were removed and saved for counting. The beads were washed several times with TE. Results are shown in Table 8.

Please replace the paragraph on page 54, lines 20-26 with the following:

A procedure for making complex arrays by PCR. A slightly complex, but considerably improved scheme to test the generality of the new approach to SBH, without the need to synthesize, separately, all 1024 five-mer probes has been developed. This procedure allows one to generate arrays with 5'- and/or 3'-overhangs and uses PCR to prepare the final probes used for hybridization which may easily be labeled with biotin. It also builds in a way of learning part or even all of the identity of each probe sequence.

Please replace the paragraph on page 55, lines 4-8 with the following:

Next, enzymatic extension of the appropriate primers using a DNA polymerase in the presence of high concentrations of dNTPs was used to make the complementary duplexes. In the above sequences, N represents an equimolar mixture of all 4 bases; R is an equimolar mixture of A and G; and Y is an equimolar mixture of T and C. The underlined sequences are *Bst XI* and *Hga I* recognition sites.

Please replace the paragraph on page 55, lines 19-22 with the following:

The sequences were designed with these internal *Bst XI*-cutting sites which allows for the generation of complementary, 4 base 3'-overhanging single-strands which can be converted to 5 base 3'-overhangs (see below) used for the type of positional SBH shown in FIG. 2A.

Please replace the paragraph on page 55, lines 25-28 with the following:

The *Hga I*-cutting site overlaps with the *Bst XI*-cutting site and allows for the generation of 5 base 5'-overhanging single-strands. This is the structure needed for the type of positional SBH shown in FIG. 2B, and can also be used for subsequent sequencing of the overhangs by primer extension.

Please replace the paragraph on page 56, lines 1-3 with the following:

The 5′- and 3′-terminal sequences of strand (a) are also recognition sites for *Sal I* and *Nhe I*, respectively; the corresponding sequences in strand (b) are recognition sites for *Xho I* and *Xma I*, respectively:

Please replace the paragraph on page 56, lines 12-26 with the following:

Those cloning sites are chosen such that, even with the degeneracy allowed by the sequences 5′-YNNNNR-3′ and 5′-RNNNNY-3′, these enzymes will not cleave the probe regions. For cloning, duplexes (a) were cleaved with both *Sal I* and *Nhe I* restriction enzymes or duplexes (b) with *Xho I* and *Xma I*. The resulting digestion products were directionally cloned into an appropriate vector (e.g., plasmid, phage, etc.), suitable cells were transformed with the vector, and colonies plated. Individual clones were picked and their DNA amplified by PCR using vector sequences downstream and upstream from the cloned sequences as the primers. This was done to increase the length of the PCR products to ease the manipulation of these products. The probe regions from individual clones were amplified by PCR with one biotinylated primer corresponding to the 5′-bases of the bottom strand. In a separate PCR, the locations of the biotins were reversed. The resulting PCR products in each case were cleaved with *Bst XI*, and the biotin-labeled products captured on streptavidin beads or surfaces. Note that by using PCR amplification instead of DNA purification, the need to separately purify and biotinylate each clone is also eliminated.

Please replace the paragraph on page 56, line 27 through page 57, line 9 with the following:

In parallel, all the PCR products were cleaved by *Hga I* which generates 5′-overhangs consisting of randomized sequences. The identity of each clone can then be determined by separate primer extensions of each of the two DNA pieces resulting from *Hga I* cleavage. For each pair of sequences, which derive from the same clone, the overhangs must be complementary. Therefore, sequencing just three bases on each fragment strand will give the entire structure of two probes. This plus/minus sequencing can be done in microtitre